Column	Comments
ACTIV Assay Status	This will indicate the status of ACTIV directed experimental investigations. As the data to populate this field do not exist yet, we would appreciate input on how best to represent this information. For example, are categories such as "none started," "ongoing," and "completed" sufficient? Would some indication of the kinds of experiments underway or completed be useful? If so, should they be split-out as distinct columns?
CDC VOC	This is meant to highlight lineages identified as a Variant of Concern (VOC) by the CDC; if there are additional outside sources that should be highlighted for their prioritization strategies would appreciate being directed to them so we can working on including them in the report.
Defining Mutations	This is meant to indicate defining mutations curated from literature for each of the lineages reported on. In the column Definitions sheet, alternative names for SARS-CoV-2 protein are also indicated. If a different format would be preferable, we'd appreciate the feedback.
Did not meet Quality or Inclusion Criteria	This is meant to indicate the number of records that we did not process either because they are from a sequencing technology we do not currently support or because the data was not of sufficient quality to generate useable results. We do plan to support additional platforms beyond Illumina, notably PacBio, in the near future. Would it be useful to separate out this causes of processing failure?
Global-USA, Doubling Time (Months)	This is meant to represent one approach to forecasting the number of records expected for a given lineage. In contrast to "Global-USA, New Records Expected this Calendar Month," here all of the historical data is considered, and the time needed to double the current total is reported. While we believe there may be better approaches to this sort of epidemiological modeling, and we are reaching out to epidemiologist for input, given the limitations in data availability noted in the "Global-USA, Sequence Record Count" section, and the associated variability in metadata submitters provide, we believe this is one of the better statistics we can provide at the moment. If this isn't useful, or an alternative statistic is preferred we'd very much appreciate the feedback.
Global-USA, Doubling Time (Weeks)	This is meant to represent one approach to forecasting the number of records expected for a given lineage. In contrast to "Global-USA, New Records Expected this Calendar Week," here all of the historical data is considered, and the time needed to double the current total is reported. While we believe there may be better approaches to this sort of epidemiological modeling, and we are reaching out to epidemiologist for input, given the limitations in data availability noted in the "Global-USA, Sequence Record Count" section, and the associated variability in metadata submitters provide, we believe this is one of the better statistics we can provide at the moment. If this isn't useful, or an alternative statistic is preferred we'd very much appreciate the feedback.

Global-USA, Growth Rate (Records/Month)	This is meant to provide a different numerical representation of how quickly a lineage is growing. It takes into account both the doubling time, "Global-USA, Doubling Time (Months)," and the number of records currently held, "Global-USA, Sequence Record Count." We'd appreciate hearing if this is adding value.
Global-USA, Growth Rate (Records/Week)	This is meant to provide a different numerical representation of how quickly a lineage is growing. It takes into account both the doubling time, "Global-USA, Doubling Time (Weeks)," and the number of records currently held, "Global-USA, Sequence Record Count." We'd appreciate hearing if this is adding value.
Global-USA, New Records Expected this Calendar Month	This is meant to represent one approach to forecasting the number of records expected for a given lineage. Simply, we look at the change over the previous two months and calculate how many we would expect this month is this trend continues. While we believe there may be better approaches to this sort of epidemiological modeling, and we are reaching out to epidemiologist for input, given the limitations in data availability noted in the "Global-USA, Sequence Record Count" section, and the associated variability in metadata submitters provide, we believe this is one of the better statistics we can provide at the moment. If this isn't useful, or an alternative statistic is preferred we'd very much appreciate the feedback.
Global-USA, New Records Expected this Calendar Week	This is meant to represent one approach to forecasting the number of records expected for a given lineage. Simply, we look at the change over the previous two weeks and calculate how many we would expect this week is this trend continues. While we believe there may be better approaches to this sort of epidemiological modeling, and we are reaching out to epidemiologist for input, given the limitations in data availability noted in the "Global-USA, Sequence Record Count" section, and the associated variability in metadata submitters provide, we believe this is one of the better statistics we can provide at the moment. If this isn't useful, or an alternative statistic is preferred we'd very much appreciate the feedback.
Global-USA, Percent	This is another view of the same information in "Global-USA, Sequence Record Count," but account for the total number of records we have processed. We believe this is especially helpful when comparing global and USA-specific trends. Are both the absolute count and the percent useful, or would only one suffice?
Global-USA, Percent New Records Released this Calendar Month to Date	Similar to "Global-USA, Percent Records Released Last Calendar Month," this is meant to provide context for interpreting data trends, and again we welcome feedback on if this is useful or if alternative or additional contextual information would be useful. Also, note the issues around data availability indicated in the "Global-USA, Sequence Record Count" section.
Global-USA, Percent New Records Released this Calendar Week to Date	Similar to "Global-USA, Percent Records Released Last Calendar Week," this is meant to provide context for interpreting data trends, and again we welcome feedback on if this is useful or if alternative or additional

	contextual information would be useful. Also, note the issues around data availability indicated in the "Global-USA, Sequence Record Count" section.
Global-USA, Percent Records Released Last Calendar Month	This is meant to provide context when interpreting statistics related to data trends by providing the number or not-USA records released the prior calendar month. If this is not useful, or if an additional piece of contextual information would be useful, we welcome the feedback. Also, note the issues around data availability indicated in the "Global-USA, Sequence Record Count" section.
Global-USA, Percent Records Released Last Calendar Week	This is meant to provide context when interpreting statistics related to data trends by providing the number or not-USA records released the prior calendar week. If this is not useful, or if an additional piece of contextual information would be useful, we welcome the feedback. Also, note the issues around data availability indicated in the "Global-USA, Sequence Record Count" section.
Global-USA, Sequence Record Count	This is meant to indicate the total number of records in NCBI's sequence databases, not from the USA, that have been associated with the indicated lineage. It is important to note that this does not include records submitted to GISAID, unless they were also submitted to NCBI. As this combines human testing, of a variety of modalities, and experimental sequencing results, this number is not easy to interpret. If a particular type of sequencing is preferred, we are happy to investigate what data we have currently and how we might be able to support reporting against it. Feedback on if the graphs at the bottom of the Executive Summary sheet are useful, especially with regards to if a particular type if graph is preferred.
Lineage	This is meant to highlight data associated lineages discussed in the scientific literature and popular media. As lineage defining mutations are typically identified by investigators other than those that maintain the nomenclature schemes, this information needs to be manually curated. Currently we include all lineages for which we can find definitions, and consult with the CDC and SPHERES consortium partners to try an ensure our definitions are aligned with what others are using. We would appreciate input on what lineages should be included, or criteria for lineage inclusion in this report, so that we can prioritize curating those lineages.
Therapeutics with Available Data	This column links to NCATS ODP website for any therapeutics tested against the associated variant lineage. We appreciate feedback on if the link in the column is sufficient, or if some other content is necessary. We'd also appreciate knowing if the linked webpage's content was what was expected or if other content was expected.
Total Records in Repository	This is meant to provide an overview of the amount of data available for downstream analyses and inclusion in other parts of the report. Would it be useful to further divide this on metadata based on the source organism, tissue types etc.?

Total Successfully Analyzed	This represent the number of records which have been successfully processed at the time of reporting, to clarify what, of all that is available, is actually included in the reporting.
USA, Doubling Time (Months)	This is meant to represent one approach to forecasting the number of records expected for a given lineage. In contrast to "USA, New Records Expected this Calendar Month," here all of the historical data is considered, and the time needed to double the current total is reported. While we believe there may be better approaches to this sort of epidemiological modeling, and we are reaching out to epidemiologist for input, given the limitations in data availability noted in the "USA, Sequence Record Count" section, and the associated variability in metadata submitters provide, we believe this is one of the better statistics we can provide at the moment. If this isn't useful, or an alternative statistic is preferred we'd very much appreciate the feedback.
USA, Doubling Time (Weeks)	This is meant to represent one approach to forecasting the number of records expected for a given lineage. In contrast to "USA, New Records Expected this Calendar Week," here all of the historical data is considered, and the time needed to double the current total is reported. While we believe there may be better approaches to this sort of epidemiological modeling, and we are reaching out to epidemiologist for input, given the limitations in data availability noted in the "USA, Sequence Record Count" section, and the associated variability in metadata submitters provide, we believe this is one of the better statistics we can provide at the moment. If this isn't useful, or an alternative statistic is preferred we'd very much appreciate the feedback.
USA, Growth Rate (Records/Month)	This is meant to provide a different numerical representation of how quickly a lineage is growing. It takes into account both the doubling time, "USA, Doubling Time (Months)," and the number of records currently held, "USA, Sequence Record Count." We'd appreciate hearing if this is adding value.
USA, Growth Rate (Records/Week)	This is meant to provide a different numerical representation of how quickly a lineage is growing. It takes into account both the doubling time, "Global-USA, Doubling Time (Months)," and the number of records currently held, "Global-USA, Sequence Record Count." We'd appreciate hearing if this is adding value.
USA, New Records Expected this Calendar Month	This is meant to represent one approach to forecasting the number of records expected for a given lineage. Simply, we look at the change over the previous two months and calculate how many we would expect this month is this trend continues. While we believe there may be better approaches to this sort of epidemiological modeling, and we are reaching out to epidemiologist for input, given the limitations in data availability noted in the "USA, Sequence Record Count" section, and the associated variability in metadata submitters provide, we believe this is one of the better statistics we can provide at the moment. If this isn't useful, or an alternative statistic is preferred we'd very much appreciate the feedback.

USA, New Records Expected this Calendar Week	This is meant to represent one approach to forecasting the number of records expected for a given lineage. Simply, we look at the change over the previous two weeks and calculate how many we would expect this week is this trend continues. While we believe there may be better approaches to this sort of epidemiological modeling, and we are reaching out to epidemiologist for input, given the limitations in data availability noted in the "USA, Sequence Record Count" section, and the associated variability in metadata submitters provide, we believe this is one of the better statistics we can provide at the moment. If this isn't useful, or an alternative statistic is preferred we'd very much appreciate the feedback.
USA, Percent	This is another view of the same information in "USA, Sequence Record Count," but account for the total number of records we have processed. We believe this is especially helpful when comparing global and USA-specific trends. Are both the absolute count and the percent useful, or would only one suffice?
USA, Percent New Records Released this Calendar Month to Date	Similar to "USA, Percent Records Released Last Calendar Month," this is meant to provide context for interpreting data trends, and again we welcome feedback on if this is useful or if alternative or additional contextual information would be useful. Also, note the issues around data availability indicated in the "USA, Sequence Record Count" section.
USA, Percent New Records Released this Calendar Week to Date	Similar to "USA, Percent Records Released Last Calendar Week," this is meant to provide context for interpreting data trends, and again we welcome feedback on if this is useful or if alternative or additional contextual information would be useful. Also, note the issues around data availability indicated in the "USA, Sequence Record Count" section.
USA, Percent Records Released Last Calendar Month	This is meant to provide context when interpreting statistics related to data trends by providing the number or USA records released the prior calendar month. If this is not useful, or if an additional piece of contextual information would be useful, we welcome the feedback. Also, note the issues around data availability indicated in the "USA, Sequence Record Count" section.
USA, Percent Records Released Last Calendar Week	This is meant to provide context when interpreting statistics related to data trends by providing the number or USA records released the prior calendar week. If this is not useful, or if an additional piece of contextual information would be useful, we welcome the feedback. Also, note the issues around data availability indicated in the "USA, Sequence Record Count" section.
USA, Sequence Record Count	This is meant to indicate the total number of records in NCBI's sequence databases, from the USA, that have been associated with the indicated lineage. It is important to note that this does not include records submitted to GISAID, unless they were also submitted to NCBI. As this combines human testing, of a variety of modalities, and experimental sequencing results, this number is not easy to interpret. If a particular type of sequencing is preferred, we are happy to investigate what data we have currently and how we might be able to support reporting against it.

	Feedback on if the graphs at the bottom of the Executive Summary sheet are useful, especially with regards to if a particular type if graph is preferred.
Variations in Other Epitopes	Similar to "Variations in Therapeutic Epitopes or Binding Sites," this field is meant to highlight lineage defining mutations in epitopes not currently associated with any therapeutics. Input on what supporting information would be most valuable is welcome.
Variations in Therapeutic Epitopes or Binding Sites	This is meant to highlight mutations in, or very near, a therapeutic-associated epitope or binding site. Currently these are combined so as not to overwhelm the reader with the number of data columns. If these should be separated out based on therapeutic class, we would welcome that feedback. In the future we hope to provide more detail on the mutations, either via a separate sheet, or, further in the future, via a link to a dedicated web resource. Input regarding what kind of supporting information would be of most value would be much appreciated.

Information provided by NCBI and NCATS for supplement to 20210412-TRACE-VariantReport-v7.xlsx